National University
of Computer & Emerging Sciences Peshawar Campus

Student Name: _____   Roll No: _____

Program: BSCS

Semester: Fall-2022

Time Allowed: 01hour

Course: _____Data Science_____

Examination: Sessional-II

Total Marks: 42 Weightage: 15

Date: 12 Nov, 2022

Instructor Name: Musadaq Mansoor

NOTE: Attempt all questions.

---

1. Draw vicious cycle and explain with a real world example? (CLO-II, Marks: 3)

2. Explain augment with an example? (CLO-II, Marks 3)

3. Explain "Build reference data" and "Archive" in ETL life cyle? (CLO-II, Marks 4)

4. Give examples of "Robust" and "Non-Robust" measures? Also explain why they are robust or non-robust? (CLO-I, Marks 4)

5. Consider having dataset of BCS (7th Semester) students, if we make the histogram of Age, What will be the modality of that histogram and why? (Consider general case, don't consider exceptional cases here) (CLO-III, Marks 3)

6. Find outliers in following data (if any) using Box-whisker method and also tell the mean , median and mode of following data (CLO-III, Marks 5)

    5,15,19,15,25,35,78,65,14,35,19,12,24,30,65,48,35

Dataset: Student.csv

Name	Program	Semester	Age	CGPA	StudyHours	GPA
John	BCS	1	19	3	2	4
Clark	BSSE	3	20	3.1	4	3
Steve	BCS	3	21	3.2	5	4
Davis	BSSE	4	20		4	2
David	BCS	3	20	2.6	7	2.5
Harvey	BSSE	8	22		4	3
Mack	BCS	7	21	3.8	6	3
Cloid	BSSE	6	22	3.9	4	4

7. For the above dataset, write PETL/Pandas/Visualisation code for the following questions keeping the below constraints in mind (CLO-IV, Marks 2*10 = 20)

Constraints:

1- CGPA is Object/String

2- Questions are not linked with one another, you can use original data for each question

Questions:

   a) Make Lineplot of StudyHours and GPA for different programs on one chart.
   b) Plot the outliers in age with respect to each semester.
   c) Make histogram of Program
   d) Make a count plot of Program

   e) Select All students having cgpa above 3

f) Select All students of BCS having age more then 20 and gpa of 4
g) Which Department has most students
h) Remove the first 3 rows
i) Change column name "StudyHours" to "No_of_Hours"
j) Apply log transformation on age